NsG-2
Michigan

MINIMUM-ERROR DEMODULATION OF BINARY PCM SIGNALS

By

Earl F. Smith

PRESENTED AT THE INTERNATIONAL TELEMETERING CONFERENCE
LONDON, ENGLAND — SEPTEMBER 23 THRU 27, 1963

RADIATION INCORPORATED

MINIMUM-ERROR DEMODULATION OF BINARY PCM SIGNALS

Earl F. Smith
Advanced Communications Group
Radiation Incorporated
Melbourne, Florida

In demodulation, as in filtering and other operations on noise-contaminated signals, a question of prime concern when contemplating improvements in the operation is, "How much improvement is theoretically possible?" Some measure of the "goodness" of the operation (e.g., rms error, error probability, etc.) must be selected if a quantitative answer to this question is desired. For any selected measure of goodness (or "performance parameter") there is generally a fundamental limitation on how well the operation can be performed with given signal power and noise conditions. Efforts to determine these limitations have led to the development of theories of optimum operations such as the Wiener filter theory (Reference 1) and theories regarding optimum detection of signals in noise (see, for example, Reference 2). Such theoretical treatments provide an optimum system (based on optimizing some performance parameter such as rms error or probability of error) whose performance may be calculated and compared with existing practical systems in order to determine the improvement theoretically possible.

In Rauch's report on improved demodulation (Reference 3), the maximum-likelihood demodulator is derived, assuming gaussian data and additive gaussian noise, for a large class of modulation operators (e.g., FM, PM, AM, PAM, PDM, etc.). However, as pointed out in that report, the results are not applicable to PCM (Pulse Code Modulation) even when the distributions of the data and the noise random processes can be assumed gaussian. Yet it appears that the maximum-likelihood criterion for demodulation should be a very meaningful one for PCM communications since maximizing the probability of selecting correct transmitted signals at the receiver is equivalent to minimizing the probability of error. However, this does not in general result in minimizing an error amplitude parameter or cost functional such as mean square error of the quantity represented by the code words.

This paper treats the PCM demodulation problem by deriving the minimum-error demodulator, assuming independent additive noise. The error probabilities (vs. signal-to-noise ratios) attained by this demodulator for band-limited white gaussian noise are then calculated. We thus obtain the lowest possible error probabilities attainable under the assumed conditions, which we can compare with error probabilities obtained with existing or proposed practical demodulation schemes.

The minimum-error demodulation of binary PCM signals with no dependence between bits and with additive, independent, band-limited, white gaussian noise has been considered in Reference 4. The term "band-limited white" here means that the power spectrum is of uniform height from zero frequency to some arbitrarily large but finite frequency, and of zero height for all higher frequencies. Comparisons are made in Reference 4 between power requirements (for any given bit-error probability) for various types of PCM transmission (e.g., PCM/AM, PCM/PSK, etc.). But the treatment emphasizes PCM/FM because of the availability of measured results from independent laboratories for PCM/FM using conventional demodulation equipment. The result of primary interest here is shown in Figure 1, in which β^2 is the ratio of signal power to noise power in a bandwidth equal to the bit rate. Comparison of the measured results with the theoretical optimum indicates that if no use is made of inter-bit dependence there are no PCM/FM demodulation schemes which will obtain specified bit-error probabilities (less than about 0.05) with significantly less transmitted power than that required by conventional demodulation equipment. Measurements of bit-error probabilities for other types of PCM transmission (PCM/AM, etc.) would very likely yield similar conclusions.

To obtain these results it was assumed that $y(t)$ during any bit interval was independent of $y(t)$ during any other bit interval. But in many cases there may be considerable statistical dependence between words (or coded samples of data) one or more frames apart. In this case, the probability distribution of the "possible" PCM signals in an interval T is not constant as assumed in Reference 4, but is higher for waveforms which exhibit this "periodic dependence" than for waveforms which do not. Hence the optimum demodulation for completely random PCM waveforms is not necessarily optimum for these more realistic waveforms, and considerable improvement in sensitivity might be gained by taking advantage of this statistical dependence.

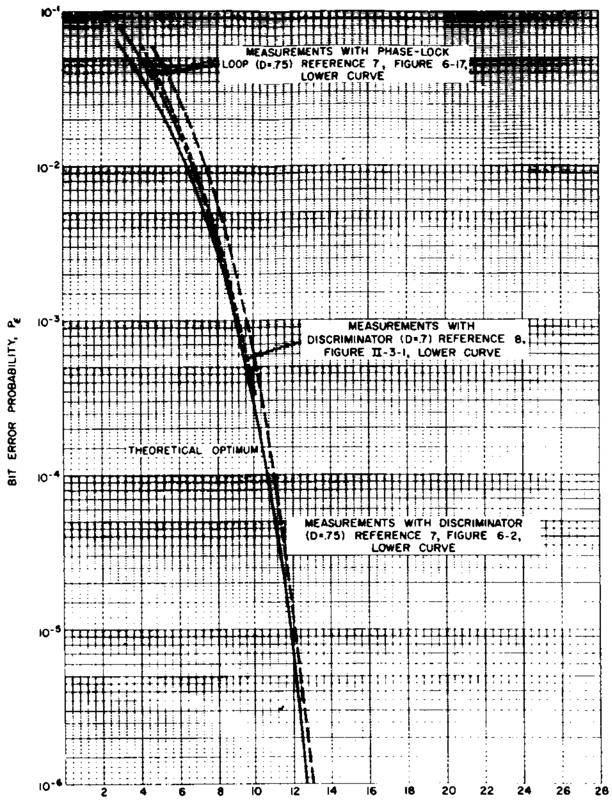

Figure 1. Bit Error Probability VS

For data power spectra which extend from zero frequency to some finite frequency, f_l, and are zero for all higher frequencies we theoretically can sample the data at $2f_l$ samples per second and recover it with zero interpolation error. The data samples in this case will be uncorrelated if the spectrum amplitude is constant from zero frequency to f_l. But for a non-idealized data spectrum whose amplitude decreases less abruptly for higher frequencies (e.g., inversely proportional to some power of the frequency) any finite sampling rate results in a non-zero correlation between samples (see Reference 5). By assuming various such non-idealized data spectra we find that if we wish to interpolate with reasonable interpolation error (say 1% rms error or less) we must sample sufficiently rapidly so that the correlation coefficients between adjacent samples may be 0.95 or greater. Although little is known about the spectral characteristics of typical measured or telemetered data, the possibility of such high correlation between data samples raises the prospect of making use of this correlation in the demodulation of telemetered data. If the correlation coefficient between two samples were unity (i.e., the probability of the two samples being identical is one) we could, by using both samples, obtain a specified signal-to-noise ratio out of the demodulator (and hence a specified error probability) with half the signal power required by using only one sample since we could, by correlation detection, effectively smooth the noise for twice as long. We now investigate the power gains made possible by correlation coefficients less than unity.

In treating minimum-error demodulation of PCM which have inter-bit dependence we treat, for simplicity, a single data source. The extension of the results to multiple data sources will be apparent. The transmitted signal, $y(t)$, is the serial binary PCM code waveform, existing over a time interval, T, representing the amplitude of successive samples of the data. The received signal, $z(t)$, is the sum of $y(t)$ and channel noise $n(t)$.

For deriving a minimum-error demodulator for PCM, it is convenient to consider the minimum error receiver to be that which determines the most likely PCM signal rather than the most likely data signal. This is a reasonable approach since there is a unique (though non-analytic) correspondence between a PCM signal and the data samples which it represents. That is, the coding and modulation operations are deterministic and reversible. Hence, determining the most likely PCM signal is equivalent to determining the most likely sequence of quantized data samples, and the interpolation of the sample values need not be considered as part of the demodulation process. Hence, maximum-likelihood demodulation may be regarded as the process of deciding which of all possible signals extending over some time interval, T, is most likely to have been transmitted, given a received noisy signal. This requires signal probability distributions for entire signals (elements of a random process ensemble).

If we consider each transmitter y-waveform extending over time interval T to be a sequence of binary PCM words, $y_1, y_2, ---, y_n$, and each received z-waveform to be a sequence of noisy binary PCM words, $z_1, z_2, ---, z_n$, then the inverse probability distribution of the transmitted waveforms may be written:

$$q(y_1, ---, y_n \mid z_1, ---, z_n) = \frac{r(z_1, ---, z_n \mid y_1, ---, y_n) g(y_1, ---, y_n)}{s(z_1, ---, z_n)}$$

where $q(y_1, ---, y_n \mid z_1, ---, z_n)$ and $g(y_1, ---, y_n)$ are probability distributions of the y-waveforms, and $r(z_1, ---, z_n \mid y_1, ---, y_n)$ and $s(z_1, ---, z_n)$ are probability density functions of the z-waveforms.

Since the a priori joint probability distribution $g(y_1, ---, y_n)$ of the word waveforms $y_1, ---, y_n$ is equivalent to the a priori joint probability distribution $f(Y_1, ---, Y_n)$ of the corresponding quantized data samples $Y_1, ---, Y_n$ (we assume $f(Y_1, ---, Y_n)$ to be known), and since $r(z_1, ---, z_n \mid y_1, ---, y_n)$ could be obtained from the assumed distribution, $h(n)$, of the additive, independent

band-limited white gaussian noise, we could evaluate $q(y_1, ---, y_n \mid z_1, ---, z_n)$ for each possible set of y_i's with a given set of z_i's, and choose the set of y_i's which gives the greatest value. But this process would not necessarily yield the lowest possible word error probability since the most probable sequence of words is not necessarily the sequence of most probable words in each word-position. If we wish to minimize word error probability we should choose the most probable word in each word-position (or "frame"). The operation which accomplishes the latter is "minimum-error demodulation".

Let us assume that we have stored n frames of received signal, and let y_i be the signal (word) transmitted in the i^{th} frame and z_i the received signal in the i^{th} frame. We wish to make a maximum probability estimate of y_i, knowing $z_1, ---, z_i, ---, z_n$. That is we wish to choose the y_i waveform which maximizes $p(y_i \mid z_1, ---, z_n)$. We may obtain $p(y_i \mid z_1, ---, z_n)$ by summing the joint probability distribution $q(y_1, ---, y_n \mid z_1, ---, z_n)$ over all y_i except y_i:

$$p(y_i \mid z_1, ---, z_n) = \sum_{y_1 \epsilon U} --- \sum_{y_{i-1} \epsilon U} \sum_{y_{i+1} \epsilon U} --- \sum_{y_n \epsilon U} q(y_1, ---, y_n \mid z_1, --- z_n) =$$

$$\sum_{y_1 \epsilon U} --- \sum_{y_{i-1} \epsilon U} \sum_{y_{i+1} \epsilon U} \frac{r(z_1, ---, z_n \mid y_1, ---, y_n) g(y_1, --- y_n)}{s(z_1, ---, z_n)} \tag{1}$$

where U = the set of all possible transmitted PCM waveforms during one word-time, T_W. The y's and z's now represent waveforms of duration T_W. If the noise is independent of the transmitted signal y then:

$$r(z_1, ---, z_n \mid y_1, ---, y_n) = h(z_1 - y_1, ---, z_n - y_n).$$

If the noise is also independent from frame to frame:

$$r(z_1, ---, z_n \mid y_1, ---, y_n) = h(z_1 - y_1) --- h(z_n - y_n). \tag{2}$$

Here h is used to represent both joint and marginal distributions for the noise random process. Then:

$$p(y_i \mid z_1, ---, z_n) = K_1(z) \sum_{y_1 \epsilon U} --- \sum_{y_{i-1} \epsilon U} \sum_{y_{i+1} \epsilon U} --- \sum_{y_n \epsilon U} h(z_1 - y_1) h(z_2 - y_2) --- h(z_n - y_n) g(y_1, ---, y_n) \tag{3}$$

where $K_1(z)$ is constant with respect to y_i.

Since $g(y_1, ---, y_n)$ is equivalent to $f(Y_1, ---, Y_n)$, equation (3) expresses $p(y_i \mid z_1, ---, z_n)$ in terms of known functions of the received z's and all possible combinations of transmitted y's. Therefore, in principle the problem of computing $p(y_i \mid z_1, ---, z_n)$ for any specified $z_1, ---, z_n$ is solved.

However, in practice the problem still appears quite formidable because of the number of operations required for the computation with a reasonable number, m, of bits per word and a reasonable number, n, of received words to be considered.

The significance of equation (3) lies in the fact that if the estimated y_i is that which maximizes (3), the probability of being wrong (i.e., probability of error) is the lowest obtainable by any method of estimation which makes use of only the n received words, $z_1, ---, z_n$. If we can calculate this error probability as a function of signal-to-noise power ratio for any assumed n and data statistics we have calculated the lowest possible error probability attainable for the assumed conditions. No explicit expression for this error probability has been obtained. Consequently, calculation of the error probability must be accomplished by a model-sampling or "monte-carlo" technique, which in its simplest form would consist of selecting sets of z's (noisy waveforms) from the proper distribution and operating on them as indicated by equation (3) to select the y_i which maximizes (3). This must be repeated, observing the frequency with which errors are made in selecting y_i, until an estimate can be made, with reasonable confidence, of the average error rate or error probability. The means for implementing this operation in a high speed digital computer must now be considered.

The expression (3) has a numerical (probability) value for any set of waveforms $y_i, z_1, ---, z_n$. In order to determine these numerical values we must first be able to determine numerical values for the factors of the form $h(z_i - y_i)$ for any waveforms z_i and y_i. These factors are the values of the noise probability density function, h(n), for $n = z_i - y_i$. We assume the noise waveform to be a band-limited white gaussian waveform of duration T_W, in which case it can be shown (see Reference 6) that if all the transmitted y-waveforms have equal energy then:

$$h(z_i - y_i) = K_2(z) \exp \left(\frac{2}{K^2} \int_{T_W} z_i y_i \, dt \right) \tag{4}$$

where K^2 is the one sided spectral height of the noise and $K_2(z)$ is constant with respect to y.

The integration is over the i^{th} word-time.

The multiple summation of equation (3) is taken over all $y_1 \in U$, $y_2 \in U$, etc., for all y_j except y_i. For m-bit words, the set U is made up of 2^m different PCM waveforms. Each of these waveforms will be distinguished in our y-notation by a second subscript. For example, the p^{th} waveform, from some ordered arrangement of the waveforms, is $y_{i(p)}$.

Furthermore, the value of the first bit of $y_{i(p)}$ will be represented by $y_{i(p)1}$, the second bit by $y_{i(p)2}$, etc. Then the exponential in equation (4) is:

$$\frac{2}{K^2}\int_{T_W} z_i y_{i(p)} dt = \sum_{r=1}^{m} \left(\frac{2}{K^2}\int_{T_B} z_{ir} y_{i(p)r} dt \right) \tag{5}$$

Where the integration of the r^{th} term of the summation is taken over the r^{th} bit-time (of duration T_B). We assume here that the bandwidth, W, of the band-limited white noise is large compared to $1/T_B$.

Let the binary character of the transmitted PCM signals be represented during each bit-time by either of two known waveforms, $g_1(t)$ and $g_2(t)$ (e.g., $g_1(t)$ representing "yes" bits and $g_2(t)$ representing "no" bits). In the treatment following it is assumed that the pair of waveforms, $g_1(t)$ and $g_2(t)$, is known for each bit-time independently of the waveform existing during any other bit-time. We must require $g_1(t)$ and $g_2(t)$ to have equal energies since we have already make that assumption implicitly by assuming that all $y_i \in U$ have the same energy. That is, we must require that

$$\int_{T_B} g_1(t)^2 dt = \int_{T_B} g_2(t)^2 dt = E_g = S^2 T_B = \frac{S^2}{B} \tag{6}$$

where S^2 = average transmitted signal power. (We will later extend our results to include any two waveforms $f_1(t)$ and $f_2(t)$ without the equal energy requirement).

The finite-time correlation coefficient for $g_1(t)$ and $g_2(t)$ is:

$$\lambda = \frac{1}{E_g}\int_{T_B} g_1(t) g_2(t) dt \tag{7}$$

It can have any value from -1 to 1.

Let $n_r(t)$ be the gaussian noise waveform during the r^{th} bit time. That is, $n_r(t)$ is the noise waveform which is added to the transmitted signal, $y_{i(s)r}$, to give the received signal z_{ir} during the r^{th} bit time. Now consider the integral

$$N_r = \frac{2}{K^2}\int_{T_B} n_r(t) y_{i(p)r}(t)\, dt$$

where $y_{i(p)r}(t)$ is either $g_1(t)$ or $g_2(t)$.

For a specific $n_r(t)$ waveform, N_r will have one value, N_{r1}, for $y_{i(p)r} = g_1(t)$ and (in general) another value, N_{r2}, for $Y_{i(p)r} = g_2(t)$. N_{r1} and N_{r2} are random gaussian variables of zero mean. By direct calculation we find their variance $\overline{N_r^2} = \frac{2S^2}{K^2 B}$ and their correlation coefficient $\frac{\overline{N_{r1} N_{r2}}}{\overline{N_r 2}} = \lambda$. Hence, if $y_{i(s)}$ is the transmitted waveform during the i^{th} word-time, the r^{th} term of the summation of equation (5) may be written:

$$\frac{2}{K^2}\int_{T_B} z_{ir} y_{i(p)r} dt = \frac{2}{K^2}\int_{T_B} \left[y_{i(s)r}(t) + n_r(t) \right] y_{i(p)r}(t) dt$$

$$= \begin{cases} \frac{2S^2}{K^2 B} + N_r & \text{for matched bits (i.e., } y_{i(s)r} = y_{i(p)r}) \\ \lambda\frac{2S^2}{K^2 B} + N_r & \text{for unmatched bits} \end{cases} \tag{8}$$

where: $N_r = N_{r1}$ if $y_{i(p)r} = g_1(t)$ (e.g., a "yes" bit)

$N_r = N_{r2}$ if $y_{i(p)r} = g_2(t)$ (e.g., a "no" bit)

Now the actual minimum error demodulator, having available only the received noisy waveforms, z_i, would determine the $h(z_i - y_{i(p)})$ values by correlating the received z_i's with each possible transmitted waveform $y_{i(p)}$ and then exponentiating the results as indicated in (4). But for purposes of simulating the operation in a digital computer (e.g., for a monte carlo method) where we must manipulate digital quantities rather than waveforms, the transmitted signals (y's) must be generated by the computer and therefore it has the information needed to determine for any assumed waveform, y_i, the bits which "match" the actual transmitted waveform, $y_{i(s)}$, and those which are "unmatched". Therefore the computer can calculate the $h(z_i - y_{i(p)})$ values of equation (4) by use of (5) and (8). Hence, the only use that the computer need make of the noise portion of the received signal waveform is to determine N_{r1} and N_{r2}. Since N_{r1} and N_{r2} are simply two correlated random gaussian variates with variance and correlation coefficient determined by $\frac{S^2}{K^2 B}$ and λ, we can, for computer simulation purposes, select N_{r1} and N_{r2} for each bit-time directly from the proper 2-dimensional gaussian distribution and use them in (8).

Note that the only characteristics of the waveforms $g_1(t)$ and $g_2(t)$ used in this operation are the mean square values, S^2, (assumed the same for both waveforms) and the finite-time correlation coefficient, λ. Hence, re-

sults obtained for any pair of waveforms apply directly to any other pair (with equal energies) having the same λ .

The basic general computer procedure then for generating sets of z's and evaluating the factors of the form $h(z_i-y_i(p))$ for a sample calculation is to first select n data samples having the appropriate correlation between samples (the means of accomplishing this is not presented here), code these samples in m-bit binary code and store these. For each code bit, select two random numbers, N_{r1} and N_{r2}, from a 2-dimensional gaussian distribution with variance $\frac{2S^2}{K^2B}$ and correlation coefficient λ (determined from the assumed $g_1(t)$ and $g_2(t)$). Evaluate each $h(z_i-y_{i(p)})$ by use of (8), (5), and (4).

For evaluating the factors of (3) of the forms $g(y, ---, y_n)$, or equivalently $f(Y_1, ---, Y_n)$, we assume the data to be from a gaussian random process with specified power spectrum (or autocorrelation function), mean m_y and variance σ_y^2. If the data quantization intervals are such that the probability density function of the data amplitude does not change appreciably over the quantization intervals, then for calculation of the joint probabilities of data samples we may use the data value at the center of a quantization interval for any data sample falling in that interval (see Figure 2). But an m-bit binary code can represent only 2^m distinct levels or quantization intervals. Hence, only a finite range of data amplitude can be represented by such a code. In Figure 2 (shown for m = 6) this finite range has been chosen to be 5.2 σ_y (or $m_y \pm 2.6\ \sigma_y$). The probability of the data occurring outside this amplitude range is less than 1%. When data samples outside this range do occur they are coded as "zero" (if below this range) or "full scale" (if above this range). The joint probability distribution, $f(Y_1, ---, Y_n)$ of quantized data samples $Y_1, ---, Y_n$ is then essentially an n^{th} order gaussian distribution with correlation coefficients ρ_{hi} equal to the values of the normalized autocorrelation function of the data, $\rho(\tau_{hi})$, where τ_{hi} is the time between samples Y_h and Y_i.

We now have all the formulations and assumptions necessary to make error probability calculations with a digital computer using the model-sampling (or monte-carlo) technique. But the computing time required increases very rapidly as m and n get large. We can learn a great deal, however, about the gains available by use of data correlation from consideration of the gain for n = 2 — that is, we examine two received (correlated) words in estimating the transmitted waveform for one of them. The cases investigated for this report are m = 6 and 3, n = 1 and 2. For m = 6, n = 2, and λ = -1, equation (3) may be written (using (4), (5), and (8)):

$$p(y_{2(p)} \mid z_1, z_2) = K_3(z) \cdot \exp\left(\sum_{r=1}^{6} \eta_{2(p)r}\, \zeta_{2r}\right) \cdot \sum_{q=0}^{63} \exp\left(\sum_{r=1}^{6} \eta_{1(q)r}\, \zeta_{1r}\right) \cdot \exp\left(-\frac{X_{1(q)}^2 - 2\rho_{12}X_{1(q)}X_{2(p)} + X_{2(p)}^2}{2(1-\rho_{12}^2)}\right) \qquad (9)$$

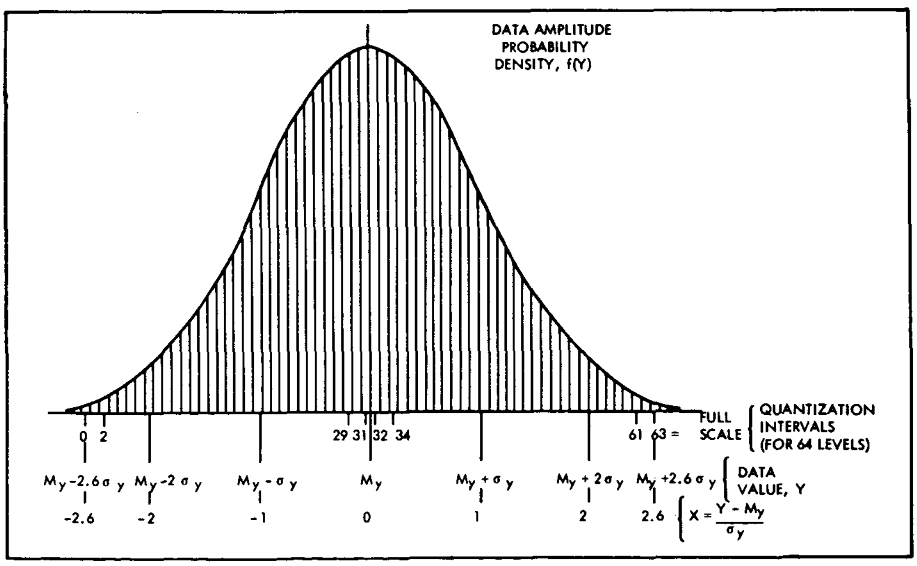

Figure 2. Assumed Data Amplitude Probability Density

Where $\eta_{i(q)r}$ is $\pm \frac{S}{K\sqrt{B/2}}$ with the algebraic sign corresponding bit of $\eta_{i(q)}$

ζ_{ir} is $\pm \frac{S}{K\sqrt{B/2}} + \nu$ with the algebriac sign determined by the corresponding bit of the transmitted word in the i^{th} word-time slot.

ν is a random gaussian variable of zero mean and unity variance.

It may be noted again that the first subscript, 1, 2, ---, j, etc., refers to the word-time-slot; the subscript in parenthesis, (p), (q), ---, refers to the particular PCM waveform; and the other subscript, r, refers to the bit-time-slot.

Since evaluation of (9) involves well defined manipulations of numbers, we may simulate the operation with a digital computer. The computations were made (using an IBM 7090 computer) for equal-energy waveforms, $g_1(t)$ and $g_2(t)$, (representing the binary character of the PCM codes) with finite time correlation coefficient, λ, of -1. These computation results can easily be used to determine the minimum attainable word-error probabilities when any two waveforms, $f_1(t)$ and $f_2(t)$, are used to represent this binary character of the codes.

Consider two arbitrary waveforms $f_1(t)$ and $f_2(t)$, of duration T_B, used for representing the binary character of the PCM waveforms. There is associated with any such pair of waveforms a "correlation" parameter, α, defined as follows:

$$\alpha = \frac{2 \int_{T_B} f_1(t) f_2(t)\, dt}{\int_{T_B} \left[f_1(t)^2 + f_2(t)^2 \right] dt} \qquad (10)$$

this parameter is identical to λ when the two waveforms have equal energies. If we subtract from both $f_1(t)$ and $f_2(t)$ the waveform $s(t) = 1/2 \left[f_1(t) + f_2(t) \right]$ we obtain two new functions, $g_1(t)$ and $g_2(t)$, which have equal energies and a λ (or α) of -1.

$$g_1(t) = \frac{1}{2}\left[f_1(t) - f_2(t) \right]$$

$$g_2(t) = -\frac{1}{2}\left[f_1(t) - f_2(t) \right]$$

This is a linear reversible operation which could be performed on the received noisy waveforms to convert the signal portions of the received waveforms from $f_1(t)$ to $g_1(t)$, but without altering the (added) noise portion of the waveforms. Then we could operate on these converted (noisy) waveforms with the minimum-error demodulator and obtain the computed error probabilities. Since the conversion operation is reversible, these error probabilities represent the minimum attainable error probabilities for the original received waveforms vs. the signal-to-noise ratio of the converted waveforms. But since the conversion does not alter the noise, the change in signal-to-noise ratio due to the conversion operation is just the square root of the ratio of the average power in the signal portions of the original and converted waveforms. We assume that "yes" bits and "no" bits (i.e., $f_1(t)$ and $f_2(t)$) occur with equal frequency since this assumption is implicit in the computed results as a consequence of assuming the data distribution of Figure 2. Then the average power of the original signal is

$$S_f^2 = \frac{1}{2} \langle f_1(t)^2 + f_2(t)^2 \rangle$$

where $\langle X \rangle$ indicates the average over one bit time, T_B, of X.

The average power of the converted signal is

$$S_g^2 = \frac{1}{2} \langle g_1(t)^2 + g_2(t)^2 \rangle =$$

$$\frac{1}{4} \langle \left[f_1(t) - f_2(t) \right]^2 \rangle =$$

$$\frac{1}{4} \langle f_1(t)^2 - 2 f_1(t) f_2(t) + f_2(t)^2 \rangle = \frac{1}{2}\left(S_f^2 - \alpha S_f^2 \right)$$

and hence:

$$\frac{\text{original signal-to-noise ratio}}{\text{converted signal-to-noise ratio}} = \left(\frac{2}{1-\alpha} \right)^{1/2}$$

Therefore the computed results may be applied to any binary PCM waveforms whatever by simply multiplying the signal-to-noise values by $\sqrt{\frac{2}{1-\alpha}}$

The computed results for n = 2, m = 6 are presented graphically in Figure 3. This figure shows the minimum-attainable word-error probability, vs. signal-to-noise ratio (normalized by $\sqrt{1-\alpha}$) when two noisy received (6-bit) binary PCM code words are used in the demodulation of one of them if the correlation coefficient between the data samples represented by the code words is ρ and the data has the gaussian amplitude distribution of Figure 2. Results are shown for ρ = 0, 0.5, 0.7, 0.9, 0.95, and 0.98. Results are also shown for completely random bits - i.e., no interbit dependence. (Note that for gaussian data there is interbit dependence due to the data amplitude distribution, even for ρ = 0).

The results of Figure 3 do not indicate as much power gain due to use of such high data correlation as might have been anticipated. This appears particularly true for word-error probabilities less than 0.1. For ρ = 0.98 the power gain is a little more than 1 db for P_W of 0.4; it is approximately 1 db for P_W of 0.2; it is less than 1 db for P_W of 0.1, and appears to continue to decrease as P_W decreases.

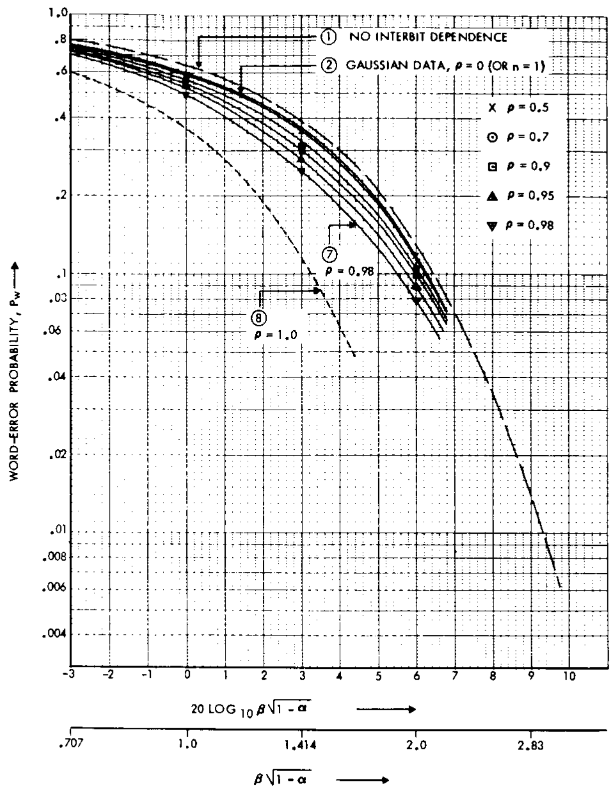

Figure 3. Word Error Probabilities vs Signal-to-Noise Ratio for n = 2, m = 6

But in determining the power gains available by use of data statistics we have assumed that the word-error probability is the "performance parameter" which is specified (i.e., fixed). If some other performance parameter such as rms error is fixed, the available gains may be different from those determined with fixed word-error probability. Of course, as discussed earlier, the (demodulation) operation should be optimized for the performance parameter of interest. But it is not unlikely that an operation optimized for one performance parameter may give considerably improved performance parameter. More specifically, the optimum demodulation operation which we have optimized for word-error probability may give considerably improved rms error performance. Some qualitative indication of this may be obtained as follows from the computed word-error probabilities for n = 2, m 3 (presented graphically in Figure 4). The results of Figure 4 for three-bit words may be interpreted as the minimum probabilities of error in the three most significant bits (i.e., the "upper half") of the six-bit words when use is made of only the "upper half" of the received words in the demodulation. But these error probabilities cannot be lower than the minimum probabilities of error in the three most significant bits when use is made of the entire received words. Hence, the minimum probabilities of error in the three most significant bits of six-bit words is equal to or less than the error probabilities obtained from Figure 4.

Therefore, for $\beta\sqrt{1-\alpha} = 2$ for example, from Figure 3 we see that a correlation coefficient of 0.98 reduces the probability of error in a six-bit word from 0.13 to 0.078 - that is, by a factor of almost 2. But from Figure 4 we see that the same correlation coefficient reduces the probability of error in the three most significant bits by a factor of 4 or more. Since errors in the most significant bits result in larger error amplitudes than do errors in the least significant bits, the above indicates that the gains afforded by use of a priori data statistics may be greater for a specified error amplitude parameter (such as rms error) than for specified error probability.

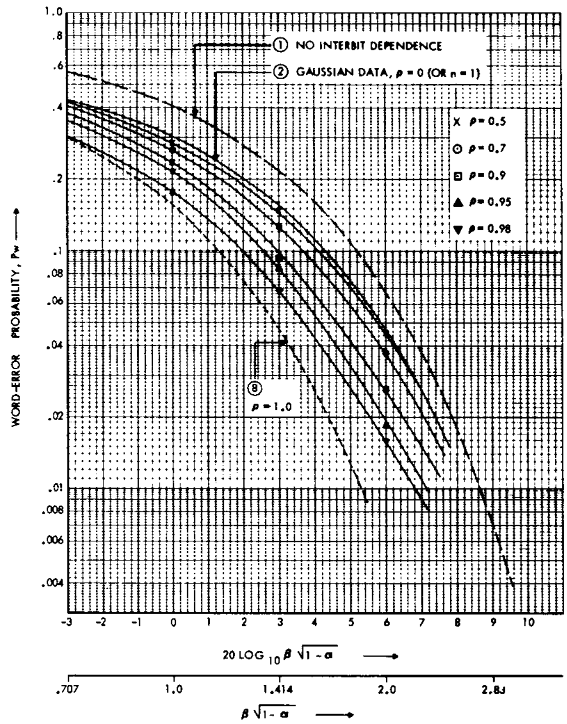

Figure 4. Word Error Probabilities vs Signal-to-Noise Ratio for n = 2, m = 3

For specified word-error probability less than about 0.1, the computed results indicate that for significant gains to accrue from the use of data redundancy, the correlation coefficients between data samples must be large (i.e., 0.98 or greater) for large numbers of samples. This apparently becomes truer as the specified word-error probability becomes lower. Whether or not sufficient redundancy is present in transmitted data must be determined by examining typical data and error requirements. There are surely cases where data sample rates are high enough so that sufficient redundancy is present, even though such high sampling rates may not be necessary for the data recovery accuracies required. In such cases the existence of the high redundancy in the data may not even be known a priori. But data redundancy which is not known a priori at the receiver cannot be used to improve the demodulation operation.

ACKNOWLEDGMENTS

The material of this paper is part of a PhD dissertation submitted to the University of Michigan. The author would like to express his gratitude to Professor L. L. Rauch of the University of Michigan for his many helpful suggestions and fruitful discussions without which this investigation would not have been started, much less completed, and for his guidance and encouragement during difficult phases. Special thanks also to Mr. D. D. McRae of Radiation Incorporated for numerous discussions which helped to clarify concepts formulated during the study, and to Dr. J. W. Young of Radiation Incorporated for his excellent help in formulating efficient logic for the computer programs.

REFERENCES

1. Wiener, N., The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications. New York: John Wiley & Sons, Inc., 1949.

2. Dwork, B. M., "Detection of a Pulse Superimposed on Fluctuation Noise". Proceedings of the IRE, Vol. 38 (July 1950), pp. 771-774.

3. Rauch, L. L., Improved Demodulation and Predetection Recording. Research Report under USAF Contract No. AF33(616)-5796, Task No. 82048, Project No. 4107, University of Michigan, 28 Feb. 1961.

4. Smith, Earl F., "Attainable Error Probabilities in Demodulation of Random PCM/FM Waveforms". IRE Transactions on Space Electronics and Telemetry, Vol. SET-8, No. 4, (Dec. 1962), pp. 290-297.

5. McRae, D. D., and Smith, E. F., Computer Interpolation. Technical Report No. 2, Parts 1 and 2, Advanced Communications Group, Radiation Incorporated, Melbourne, Florida, Part 1, 15 Dec. 1961, Part 2, 10 May 1962.

6. Woodward, P. M., Probability and Information Theory with Applications to Radar. New York: McGraw Hill Book Co., Inc., 1953.

7. Experimental Determination of Signal-to-Noise Relationships in PCM/FM and PCM/PM Transmission. Interim Report for NASA, GSFC, Contract NAS 5-505 Electro-Mechanical Research, Inc., Sarasota, Florida, 20 Oct. 1961.

8. Telemetry System Study Final Report, Volume II of III. "Experimental Evaluation Program". U. S. Army Signal Research and Development Laboratories Contract No. DA-36-039-SC-73182. Aeronutronic; Newport Beach, California, 18 Dec. 1959.